
CHAPTER IV. VISUAL ANALYSIS — A MEANS OF EXPLORATION IN STATISTICAL RESEARCH

CONTENTS

Visual analysis—A means of exploration in statistical research-----	Page 227
ILLUSTRATIONS	
Figure 1.—Three stages in the preparation of visual-analysis cards—blank form ready for transcription, transcribed card, and coded and slotted card-----	229
Figure 2.—Three cards each representing a farm and farm family. All of the information used in the study concerning the farm and the operator's family is carried on one card-----	230
Figure 3.—Pack of slotted cards before beginning exploratory studies of relationships-----	231
Figure 4.—Sorting slotted cards by means of a needle-----	232
Figure 5.—Cards sorted to show the relationships existing between size of farm and selected characteristics-----	233
Figure 6.—Card designed primarily for visual-analysis purposes-----	235
Figure 7.—Cards sorted first by size of farm, then each size group halved to illustrate the stability of the relationships visible-----	235
Figure 8.—Cards sorted to show the effect of tenure on size of farm; also, that some characteristics related to size of farm are not affected by tenure-----	236
Figure 9.—Graphic illustration of the changes in number of farms by tenure which accompany changes in the size of farms used in these examples-----	237
Figure 10.—Farms reporting automobiles compared with those not reporting automobiles. In these examples characteristics associated with large farms rarely appear on cards representing farms not reporting automobiles-----	239
Figure 11.—Graphic illustration of the proportions shown in figure 10-----	239
Figure 12.—Cards representing farms reporting trucks are shown below the heavy line-----	241
Figure 13.—Relative lengths of bars indicate the changing proportions in these examples of farms reporting trucks in six size-of-farm groups-----	241
Figure 14.—Cards sorted to separate farms reporting tractors from those not reporting tractors. Cards used in these examples show most of the smaller farms in the nontractor group-----	243
Figure 15.—Bar graph showing, for the farms represented by the cards in figure 14, the increase in proportion reporting tractors as the size of farms increases-----	243
Figure 16.—Progressive selection. Final group of cards, shown at top, represents farms reporting automobiles, tractors, and trucks, and having electricity available-----	245

VISUAL ANALYSIS—A MEANS OF EXPLORATION IN STATISTICAL RESEARCH

A representative of the research division of a national farm magazine made a visit to the Bureau of the Census for the purpose of discussing tabulation of material which would be of special value to advertisers using his magazine as a medium. The Census had a few months earlier made a subscriber study for the company and as a result had on hand all of the material previously prepared.

The representative was, of course, familiar with the methods used in making that study. His purpose in making a personal visit to the Census was to determine, as a preliminary to further tabulating, what the results would be. Having knowledge of the preparation for the earlier study, he knew that the Census was in position to furnish, in a matter of minutes, a preview of the results to be expected in any suggested tabulation of the material used. After having examined the results of several proposed tabulations, he arranged for payment of the costs and ordered selected tables prepared.

The representative was enabled to "see" the results to be accomplished because of the use in his study of a special card which provided a means of visually analyzing the material on hand. This visual analysis enabled the Census not only to present the probable results of several specified tabulations, but actually to show the representative approximately what those results would be when tabulations of his material had been made.

One of the questions asked of the Bureau of the Census was this—"Among our subscribers farms, what are some of the characteristics which distinguish those on which tractors are used as a source of farm power from those on which tractors are not used?"

Previously, no tables had been prepared which could answer this question, but within just a few minutes the information, applicable to the subscribers farms included in the sample, was supplied to the representative:

1. On most of the farms using tractors, the major source of income is from field crops; next in importance are those with major source from livestock sold or traded; and third, those with major source of income from dairy products.
2. Late model tractors were found among livestock and dairy farmers more often than among those whose major source of income was from crops.
3. Late model automobiles tended to accompany late model tractors.
4. Livestock and field crop farms reporting tractors ranged in size upward from 70 acres. Dairy farms ran smaller.
5. The majority of farms of over 140 acres reported tractors.
6. Only a very small proportion of farms of less than 70 acres reported tractors.
7. Farms without tractors reported fewer automobiles than did the farms with tractors.

With this information the representative ordered three separate tabulations which would show the numerical extent of the principal relationships brought out in this preliminary analysis.

¹The system of visual analysis described in this chapter was developed by Carl R. Nyman and Orville C. Demaree of the Division of Agriculture, Bureau of the Census. In a routine search of the Patent Office records made in the interest of the Government, expired Letters Patent, No. 1,150,793, were found issued to Charles H. Talmadge on August 17, 1915. In the letters patent was described a card embodying a system of classification by means of peripheral slots similar to the cards used in the illustrations accompanying this article.

The researcher is very often faced with the problem of determining which way to proceed. Given a wide and detailed knowledge of the subject with which to set up hypotheses the researcher can proceed great distances into uncharted fields. As the development of the project proceeds, however, problems arise. One of these is that of wasted effort. A long series of tabulations and computations completed and found somewhat short of supporting the hypothesis may prove, for example, that certain factors do not influence the final result to a significant extent. This information may be valuable, but often must be considered as wasted effort to the extent that it could have been avoided. In other words, if a shorter and more efficient means of determining the lack of influence of these factors had been used, the same results would have been known more quickly and at less cost.

Another problem arising is one involved simply with the completeness of the analysis. In proceeding with research and analysis with a given amount of data, when is the problem finished, or, when have all important factors been considered? With only three or four items or characteristics involved, the answer to this question can very often be obtained in a relatively short time. But in any analysis involving Census data when dozens of items of information bearing on certain subjects are available, the question is very definite. The selection, in advance, of all of the characteristics which will be required to complete the study, or the selection of those which have the most bearing on the problem in question, are of paramount importance. The funds available for a given project very often limit the study to the completion of a single planned attack. When, upon nearing the completion of the project, the analyst discovers other characteristics which look promising and he is unable to use them in place of some which proved of little value to his study, he is likely to feel as Moses must have felt when gazing over the land of Canaan.

What is needed then is a means of determining at low cost in time and effort the answers to these two questions: (a) What are the factors which are likely to be of influence in the study, and (b) which are of greatest importance. When the answers to these questions are known, the analyst can proceed with the knowledge that, when the project is completed, he will have thoroughly covered the subject within the limits prescribed.

With such a means at hand and used to its utmost capabilities, exact factors of influence can be determined, leaving for tabulations and further study the determination of the extent of such influence under the various restrictions imposed in the study. The purpose of this article is to describe a means of obtaining in advance of tabulations a thorough understanding of the factors involved and a general knowledge of their relative importance. The Census has been using for the past year a system which has amazing possibilities when used in many research problems. The system provides a means of keeping "visible" at all times all items involved in the study, and also a means of noting relationships existing between these items. Relationships between, and the relative importance of, the various characteristics can actually be seen on the edges of a pack of visual-analysis cards.

The principles involved in the visual-analysis system are simple. Using the individual farm as an example, all possible characteristics descriptive of each farm are coded and slotted in edges of cards. The cards may be of any convenient size and shape. On each card is carried a prepared form to which is transcribed from all sources all of the information to be used or considered in a particular study.

Figure 1 shows the three stages of development of the cards. The card at the top shows the prepared form ready for transcription, the center card is a sample of the transcribed data, and the lower card shows the same data coded and slotted in the margins of the card. Note that holes are provided for each code, but only the appropriate code is slotted for each item.

The particular cards used in this and all following illustrations were used in a special study mentioned in the opening paragraphs of this chapter. Cards designed specifically for visual analysis would be somewhat different. A layout for a strictly exploratory visual-analysis card expected to be used in the Census is shown later in figure 6.

For purposes of simplification, only a few items or characteristics of farms will be discussed. The purpose is to explain exactly how the system works rather than to work out any particular problem. Since it will be well to become familiar with these particular items in order to more readily understand the descriptions and illustrations to follow, a brief listing and a description of each of the selected items are here given.

Referring to the upper edge of the top card, at the left is a block of 7 holes labeled "Tenure." A card slotted in the first hole represents a farm operated by a full owner, or in other words, the operator owns all of the land he operates. Slotted in the second hole the card represents a farm operated by a part owner or a farm of which the operator owns a part and rents a part from others. Code 3 refers to managed farms and codes 4, 5, 6, and 8 to various forms of tenancy. Further explanation concerning tenure will not be necessary.

In the block of 6 holes labeled "Size of farm" the first hole on the left refers to farms of less than 30 acres. The second hole refers to farms of 30 to 69 acres, the third to 70 to 139 acres, the fourth to 140 to 179 acres, the fifth to 180 to 259 acres, and the last to farms of 260 acres and over. Question 12 in the upper section of the printed form on the card carries the acres in farm.

In the block marked "Mortgage" the left of the 3 holes, with the word "Yes" printed underneath, is slotted when the reply to Question 22 is "Yes" and indicates that a farm is operated by a full owner or part owner and carries a mortgage on the land and buildings of the farm. The center hole is slotted if the reply to Question 22 is "No," indicating that the farm is operated by a full owner or part owner and is free from mortgage. The hole at the right marked "Amt." is slotted when the amount of the mortgage is shown in reply to Question 23 on a full-owner or part-owner farm.

Holes are slotted in the block marked "Co-op" in accordance with replies to Questions 34, 35, and 36, a reply of "Yes" indicating that the farm operator did business with or through the kind of cooperative organization shown. The 3 types of cooperatives shown refer to selling, buying, and service cooperatives.

Holes in the blocks for "Autos," "Trucks," and "Tractors," are slotted in accordance with replies to these questions in the body of the card. In each group the hole to the left is slotted if the year of latest model is 1936 or later, the center hole is slotted if the year of latest model is 1931 to 1935, and the right hand hole is slotted if the year of latest model is earlier than 1931. In addition, to conserve space by not using an additional hole in each set, the right hand hole was slotted whenever a machine of any year model was reported. Thus, every card showing a machine reported was slotted in the right hand hole and if the year of latest model was before 1931, no other holes in the block were slotted. If the year of latest model was 1931 or later, one of the other two holes in the appropriate block was slotted.

In this discussion, only the left hand hole of the block marked "Electric" will be noted. This hole is slotted whenever the reply to Question 47 indicates that there was an electric distribution line within 1/4 mile of the operator's dwelling. The other two holes indicate the source of current when the operator's dwelling was lighted by electricity. They are "power line" and "home plant," respectively.

The hole marked "Tel" is slotted whenever the reply to Question 49 indicated that a telephone was reported.

All other holes on the card were notched in a similar manner in accordance with information shown on the card. For

the purpose of this discussion, however, a clear understanding is necessary of only those described.

After the material has been transcribed and coded and the appropriate slots made in the edges of the card, the slots on any individual card may be said to form a "pattern" for the farm represented. Another farm having the same characteristics would, therefore, have the same pattern of slots and, conversely, any difference in the pattern for individual farms would indicate corresponding differences in their characteristics. In figure 2 are shown such cards for three individual farms. Examination of these will show that as the characteristics of the farms change so do the patterns of slots.

When a group of cards, each card representing an individual farm, are placed together in a pack, the patterns of the individual farms are, of course, not visible. Only the edges are exposed. Figure 3 shows a pack of approximately 550 cards representing that number of individual farms. These particular farms are located for the most part in the Corn Belt. While many of these farms with respect to any one characteristic are similar, there are probably no two farms in the entire group which have exactly the same individual pattern of characteristics. But, as will be seen shortly, when two or more farms have one similar characteristic as indicated by slots in the same position, they are very likely to be similar in one or more other characteristics, and as a consequence have corresponding slots in similar positions. It is this coincidence of slots representing similar characteristics of farms which forms the basis for the visual system of exploration being described.

When a number of contiguous cards possess the same characteristic, this fact becomes evident in the appearance of a groove formed by the several coincident slots. Any process which places together all cards showing identically placed slots has automatically selected from the pack all cards representing farms similar in the characteristic represented by the particular slot being used in the selection. When these cards are placed in a separate pack by themselves, the slots representing this characteristic form a continuous groove across the pack. Figure 4 illustrates the manner in which slotted cards are sorted.

Note that the "needle" has been inserted entirely through the pack in holes provided in that position, and, upon being raised, all cards that have been slotted in that position fall and can be readily assembled with others having a similar slot. By simply sorting the cards, all farms possessing a specified characteristic are placed together. When the various characteristics are described by proper notation on the cards themselves, as has been done in these examples, the sorting of farms by characteristics is simple and rapid.

Earlier in this chapter it has been stated that when farms are similar with respect to one characteristic, they are likely to be similar with respect to others. In varying degrees, certain characteristics are accompanied by certain other characteristics and any change in the one is likely to be accompanied by a corresponding change in others. The closer the relationship between items, the more likely are these accompanying changes to occur. If the correlation were perfect between two items, any change in one would always be accompanied by a similar change in the other. Such perfect correlation, however, is hardly to be expected in any considerable number of cards.

When a pack of cards, sorted according to any given characteristic, is viewed from the edge, the slots representing that particular characteristic, as has been said, form a continuous groove across the pack. Slots representing other characteristics most closely related to the one used in sorting will tend to assemble in a similar group. In case of perfect correlation, another groove would be formed across the pack. While this rarely happens, it is not uncommon for as many as 90 percent of the cards representing a group of farms possessing one characteristic to also possess the same second characteristic. Less closely related characteristics will be represented by fewer continuous adjacent slots forming grooves. By this means the relative relationship between items can be judged by the relative frequency of coincidence of slots—the greater the frequency of coincidence, the closer the relationship.

Figure 1 displays three stages in the preparation of visual-analysis cards: a blank form ready for transcription, a transcribed card, and a coded and slotted card.

The top card is a blank form titled "FARM SCHEDULE". It contains various sections for data entry, including:

- 1. Name of operator**: Blank for transcription.
- 2. Address**: Blank for transcription.
- 3. Age**: Blank for transcription.
- 4. Acres owned**: Blank for transcription.
- 5. Beside on farm**: Blank for transcription.
- 6. Value of crops**: Blank for transcription.
- 7. Value of farm**: Blank for transcription.
- 8. Value of buildings**: Blank for transcription.
- 9. Value of machinery**: Blank for transcription.
- 10. Any mortgage**: Blank for transcription.
- 11. Taxes**: Blank for transcription.
- 12. On additional land**: Blank for transcription.
- 13. Work off farm**: Blank for transcription.
- 14. Cooperative**: Blank for transcription.
- 15. Labor-Sep.**: Blank for transcription.
- 16. Family**: Blank for transcription.
- 17. Expended for**: Blank for transcription.
- 18. Gas, etc.**: Blank for transcription.
- 19. Auto**: Blank for transcription.
- 20. Tractor**: Blank for transcription.
- 21. Radio**: Blank for transcription.
- 22. Telephone**: Blank for transcription.
- 23. Highest Ed.**: Blank for transcription.
- 24. Year built**: Blank for transcription.
- 25. Number of rooms**: Blank for transcription.
- 26. Fuel for heating**: Blank for transcription.

The middle card is a transcribed version of the blank form, with handwritten data entered in the corresponding fields.

The bottom card is a coded and slotted version of the transcribed card, with the data organized into a structured format for visual analysis.

FIGURE 1.—Three stages in the preparation of visual-analysis cards—blank form ready for transcription, transcribed card, and coded and slotted card.

FARM SCHEDULE										FARM SCHEDULE										FARM SCHEDULE										FARM SCHEDULE										FARM SCHEDULE									
1. Name of operator										2. Address										3. Date of birth										4. Date of death										5. Date of birth									
6. Name of operator										7. Address										8. Date of birth										9. Date of death										10. Date of birth									
1. Name of operator										2. Address										3. Date of birth										4. Date of death										5. Date of birth									
6. Name of operator										7. Address										8. Date of birth										9. Date of death										10. Date of birth									
11. Name of operator										12. Address										13. Date of birth										14. Date of death										15. Date of birth									
16. Name of operator										17. Address										18. Date of birth										19. Date of death										20. Date of birth									
21. Name of operator										22. Address										23. Date of birth										24. Date of death										25. Date of birth									
26. Name of operator										27. Address										28. Date of birth										29. Date of death										30. Date of birth									
31. Name of operator										32. Address										33. Date of birth										34. Date of death										35. Date of birth									
36. Name of operator										37. Address										38. Date of birth										39. Date of death										40. Date of birth									
41. Name of operator										42. Address										43. Date of birth										44. Date of death										45. Date of birth									
46. Name of operator										47. Address										48. Date of birth										49. Date of death										50. Date of birth									
51. Name of operator										52. Address										53. Date of birth										54. Date of death										55. Date of birth									
56. Name of operator										57. Address										58. Date of birth										59. Date of death										60. Date of birth									
61. Name of operator										62. Address										63. Date of birth										64. Date of death										65. Date of birth									
66. Name of operator										67. Address										68. Date of birth										69. Date of death										70. Date of birth									
71. Name of operator										72. Address										73. Date of birth										74. Date of death										75. Date of birth									
76. Name of operator										77. Address										78. Date of birth										79. Date of death										80. Date of birth									
81. Name of operator										82. Address										83. Date of birth										84. Date of death										85. Date of birth									
86. Name of operator										87. Address										88. Date of birth										89. Date of death										90. Date of birth									
91. Name of operator										92. Address										93. Date of birth										94. Date of death										95. Date of birth									
96. Name of operator										97. Address										98. Date of birth										99. Date of death										100. Date of birth									
101. Name of operator										102. Address										103. Date of birth										104. Date of death										105. Date of birth									
106. Name of operator										107. Address										108. Date of birth										109. Date of death										110. Date of birth									
111. Name of operator										112. Address										113. Date of birth										114. Date of death										115. Date of birth									
116. Name of operator										117. Address										118. Date of birth										119. Date of death										120. Date of birth									
121. Name of operator										122. Address										123. Date of birth										124. Date of death										125. Date of birth									
126. Name of operator										127. Address										128. Date of birth										129. Date of death										130. Date of birth									
131. Name of operator										132. Address										133. Date of birth										134. Date of death										135. Date of birth									
136. Name of operator										137. Address										138. Date of birth										139. Date of death										140. Date of birth									
141. Name of operator										142. Address										143. Date of birth										144. Date of death										145. Date of birth									
146. Name of operator										147. Address										148. Date of birth										149. Date of death										150. Date of birth									
151. Name of operator										152. Address										153. Date of birth										154. Date of death										155. Date of birth									
156. Name of operator										157. Address										158. Date of birth										159. Date of death										160. Date of birth									
161. Name of operator										162. Address										163. Date of birth										164. Date of death										165. Date of birth									
166. Name of operator										167. Address										168. Date of birth										169. Date of death										170. Date of birth									
171. Name of operator										172. Address										173. Date of birth										174. Date of death										175. Date of birth									
176. Name of operator										177. Address										178. Date of birth										179. Date of death										180. Date of birth									
181. Name of operator										182. Address										183. Date of birth										184. Date of death										185. Date of birth									
186. Name of operator										187. Address										188. Date of birth										189. Date of death										190. Date of birth									
191. Name of operator										192. Address										193. Date of birth										194. Date of death										195. Date of birth									
196. Name of operator										197. Address										198. Date of birth										199. Date of death										200. Date of birth									
201. Name of operator										202. Address										203. Date of birth										204. Date of death										205. Date of birth									
206. Name of operator										207. Address										208. Date of birth										209. Date of death										210. Date of birth									
211. Name of operator										212. Address										213. Date of birth										214. Date of death										215. Date of birth									
216. Name of operator										217. Address										218. Date of birth																													

FIGURE 2.—Three cards each representing a farm and farm family. All of the information used in the study concerning the farm and the operator's family is carried on one card.

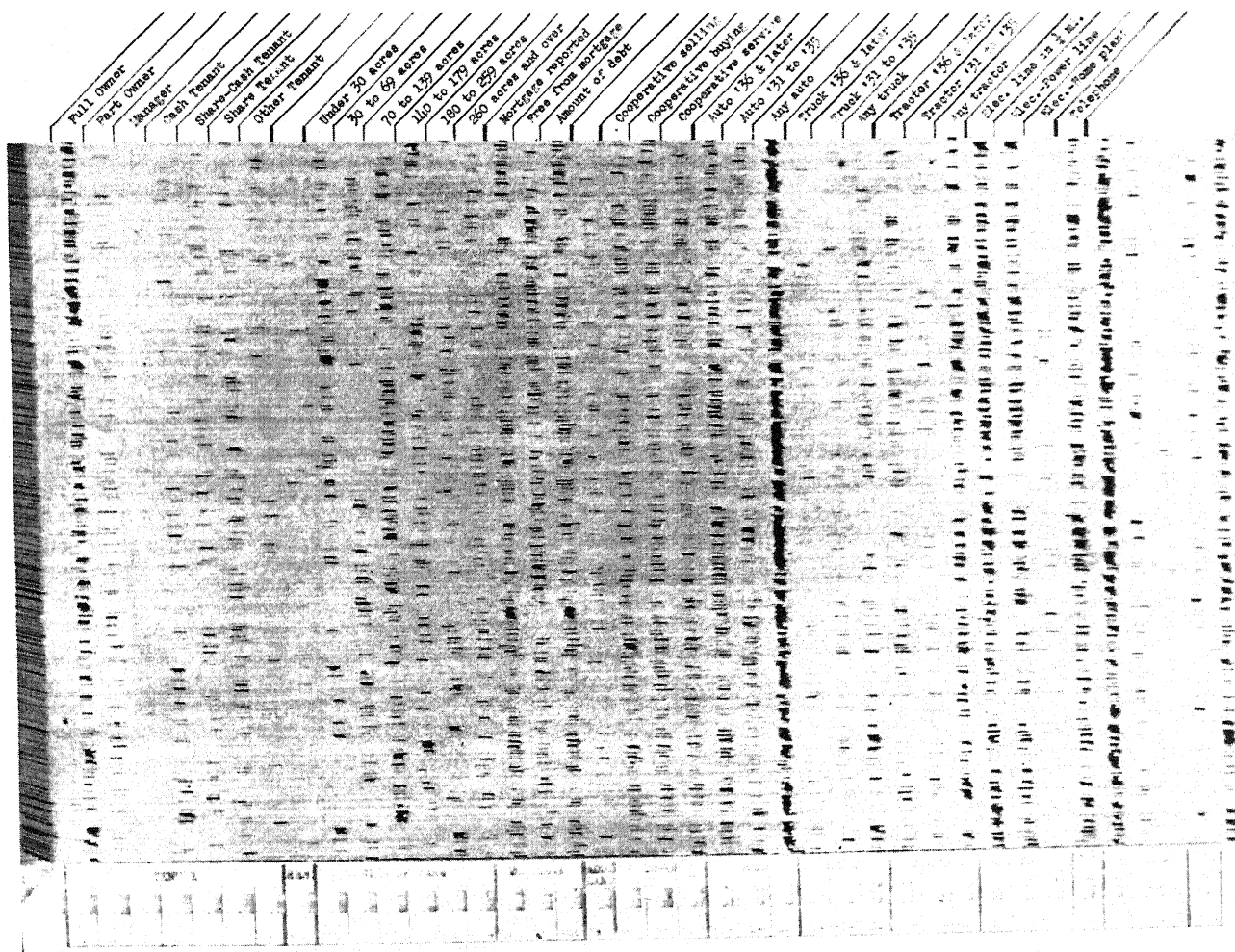


FIGURE 5.—Pack of slotted cards before beginning exploratory studies of relationships.

It will be realized, of course, that all indicated relationships between items are not, of necessity, cause and effect relationships. Changes in some items may be the effect of a causal factor not immediately evident. Especially is this likely to be true when only a limited number of characteristics are coded on the edge of the card as is the case in these examples. It is entirely possible that two or more characteristics carried on a limited card may vary in the same way and still not be influenced by changes in each other. This, however, does not detract in the least from the value of visual exploration. As nearly as possible, all characteristics should be coded and slotted and any additional characteristics suspected of having influence, but concerning which information is not available, should be kept in mind during the entire progress of the study.

Since relationships are indicated by the frequency of coinciding slots, we can speak of the relative frequencies of the slots for each coded characteristic as the pattern of characteristics for a group of farms. Hence, each time the cards are sorted, a new group is brought into existence with an accompanying new pattern of characteristics.

Before proceeding further, it is important that the reader have clearly in mind that the cards herein used for illustrative purposes are a portion of the cards used in a special study and do not represent any particular group of farms. Hence, relationships appearing in this and subsequent pictures cannot be considered as being representative of farms generally or specifically. On the other hand, the cards were not selected to show outstanding characteristics of any kind. They are strictly "run of the mill" cards and any similar group of cards will work exactly the same way.

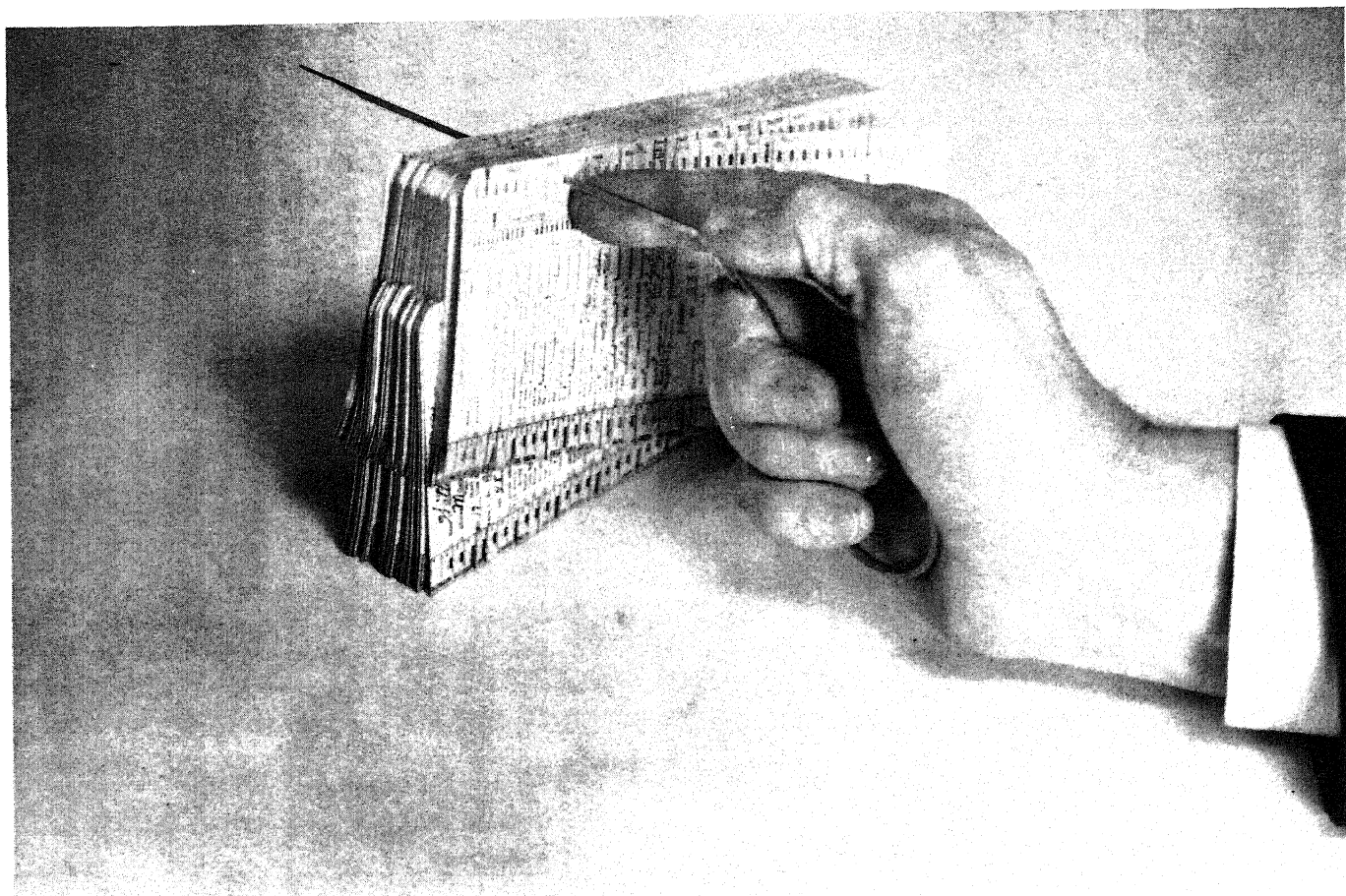


FIGURE 4.—Sorting slotted cards by means of a needle.

Figure 5 shows the pack of cards sorted by size of farm. The cards included between the limits of each continuous groove represent all farms within the limits of each corresponding interval described in the legend at the top of the picture.

Having sorted the cards once, we are now in position to note several characteristics of these farms as they relate to this sort. In other words, the relationships of the various characteristics to size of farm are now visible. The reader's attention is probably first drawn to the column indicating possession of an automobile. In this column the frequency of slots is greater than in any other and indicates that a very high percentage of the farms report automobiles. But the frequency of slots is not uniform throughout the length of the column. At the top of the column, only an occasional card is not slotted, but as the eye travels downward the unslotted cards become increasingly numerous. It is evident that a greater proportion of the operators of the smaller farms are without automobiles than is the case with the operators of the larger farms.

Tenure also changes with size of farm. In the farms of less than 70 acres (those included in the first two size bands in the lower portion of the picture) the proportion of full owners is much higher than in farms larger than 70 acres. Another very noticeable item is that there are practically no part owners operating these smaller farms. However, near the top of the part-owner column, particularly on farms of 260 acres and over, there are more part owners than full owners.

The next observation of interest is the fact that the slots in the "Yes" column of the mortgage group tend to increase in frequency with increase in size of farm when compared with the slots in the "No" column. Therefore, it can be definitely stated that on the farms represented by these cards, the proportion of mortgaged farms increases as the size of farm increases.

In the field of cooperative business, little is shown in the groups of less than 70 acres, but it is rather uniform above this point.

Proportion of later model automobiles (1936 and later) seems to increase with size of farm.

Some relationship can be seen between trucks and size of farm.

Farms reporting tractors very definitely increase with size of farm.

Farms with electricity available (distribution line within 1/4 mile of operator's dwelling) are relatively more numerous in the smaller sizes, but telephones are distributed in the opposite direction—more of the larger farms reporting telephones.

Summarizing the above observations:

1. Items which vary directly with size of farm—
 - a. Part-owner-operated farms.
 - b. Proportion of owned farms reporting mortgage.
 - c. Cooperative business.
 - d. Automobiles.
 - e. Trucks.
 - f. Tractors.
 - g. Telephones.
2. Items which vary inversely with size of farm—
 - a. Full-owner-operated farms.
 - b. Availability of electricity.

The discussion up to this point has been for the purpose of acquainting the reader with the general features of the system before including too many details as to how it may be used. The main purpose of this chapter will be to present to the reader as clearly as possible by means of illustrations and descriptions the general features of the system leaving its adaptation to particular fields of research to the imaginations of the research workers interested.

The number of cards which can be conveniently handled in any given project is limited. Probably 10,000 cards is nearing the maximum for most jobs unless the work divides itself early in the study. For exploratory work which does not include too much detail a number less than one-half this large will probably be adequate. In general, the greater the number of characteristics involved, the greater will be the number of cards required. As the detail increases the number of ramifications, the ultimate groups will become too small for stability of results if too few cards are used.

For small jobs in which the number of reports is only a few hundred or, in some cases, a few thousand, the entire study may be made from the cards without additional punching and machine tabulation. Pick-ups may be made of the transcribed data directly from the cards by means of adding machines.

For larger jobs in which the number of reports is so large as to make the cost of hand tabulations too high, a sample of the reports may be taken and explorations made by means of visual analysis. Since the greatest value of the card lies in its use for visual analysis, most of the exploration is made without tabulations of any kind and without even counting the cards. A draftsman's divider is usually all that is necessary for quickly measuring the lengths of grooves, when it is desirable to record them for future reference. Rarely will it become necessary in an exploratory study to record other than

relative lengths of grooves, at least in the early stages. Occasionally it will be desirable to count the cards showing certain characteristics and record the counts.

Not the least of the advantages of the visual-analysis cards is the fact that all of the operations can be performed by the person most interested in the work or responsible for the planning and direction of the work. When the final plans have been completed, the researcher knows from his own personal study of the cards what the results will be and that complete tabulations will furnish the statistical proof of his conclusions.

Knowing the relationship between the various characteristics of farms is a valuable aid in much of the work of the Agriculture Division of the Bureau of the Census. Presentation of results is being constantly improved through studies of relevancy, making the published data more valuable and more adaptable to the needs of the users of such data. In this connection, one of the principal advantages is the fact that after the cards have been prepared and tabulations originally planned have been completed, subsequent tabulations suggested by results of the originals can be made at comparatively low cost. The cards are drawn up in such manner that each card carries all of the information concerning a particular farm and/or family, and may be used in every conceivable combination or separation without further preparation.

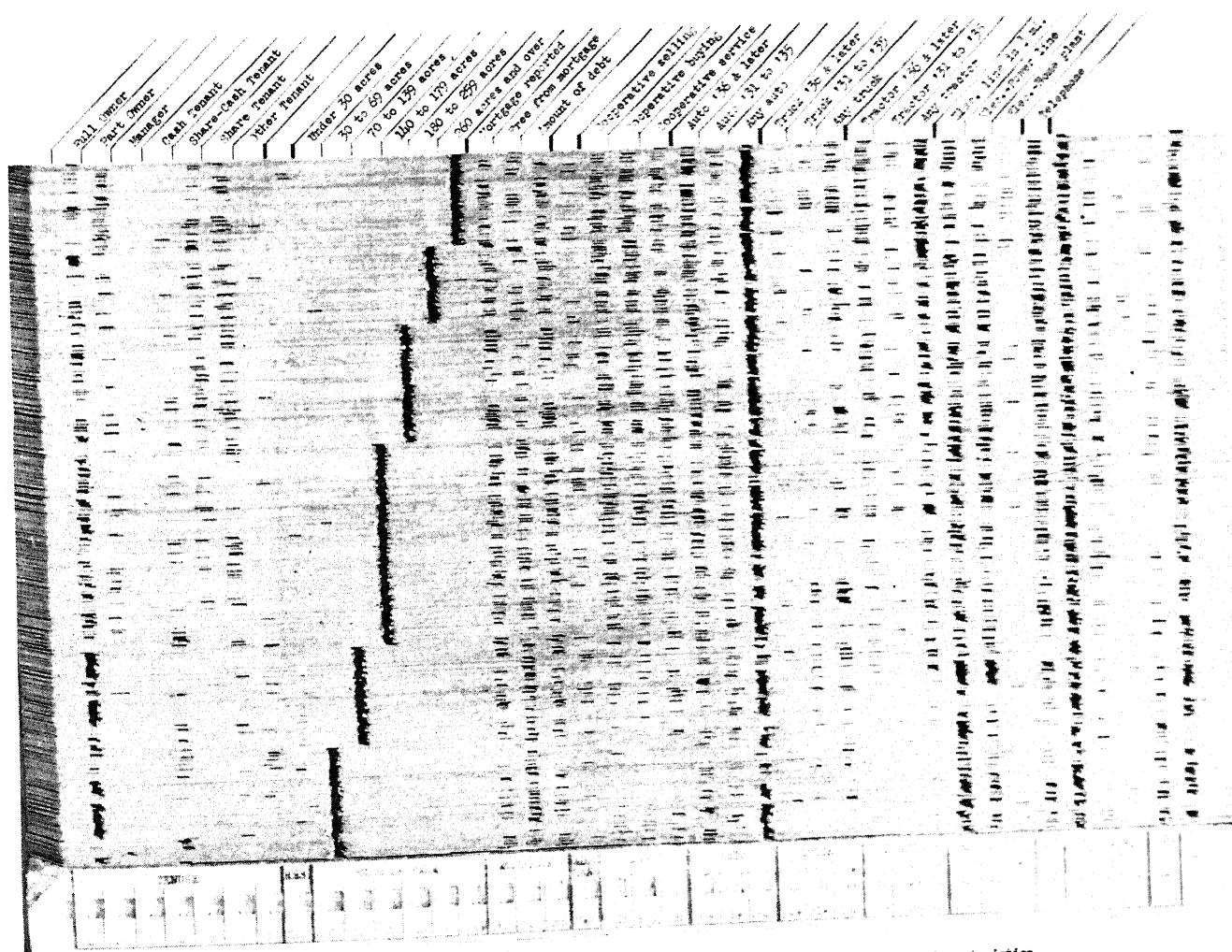


FIGURE 5.—Cards sorted to show the relationships existing between size of farm and selected characteristics.

In figure 6 is shown a layout of a card being developed for purely exploratory work in the Division of Agriculture of the Census. It is shown here merely as an example of such layout and to illustrate certain refinements not included on the cards used in the examples. It is strictly a visual-analysis card. In order to conserve space, many of the descriptive terms have been so shortened as to be unintelligible without a key unless the user is familiar with these terms. It can be seen readily, however, that provision has been made for coding and slotting on the outer edge of the card a much greater number of characteristics than was possible on the cards used in the other illustrations in this paper.

The card has two important features which distinguish it from cards designed primarily for tabulation purposes. Being a visual-analysis card, the arrangement of the bands, or frequency groups, in two lines of holes is definitely for the purpose of increasing the "visibility" of the relationships. Most of the bands in the outer row are combinations of two or more bands in the inner row. This serves to increase the group interval in frequency distributions without sacrificing the narrower groupings when a more detailed break-down becomes desirable. The wider intervals serve to avoid a scattered appearance when the range of relevancy may be within rather narrow limits. If too many bands are used in the outer row, there is a possibility that a relationship might pass unnoticed due to the fact that the slots for a particular characteristic might scatter among three groups, whereas if the three groups were combined, all would fall in the same slot. This feature is of greatest importance when the number of cards is limited. Another, and by no means small, advantage of the double line method of coding and slotting is that it greatly increases the capacity of the card.

The other feature of the card is the arrangement of the material on the body of the card designed to facilitate transcription and coding. Once these operations have been completed relatively little use is made of the figures themselves. Therefore, ease of locating and reading the transcribed items should be subordinated to ease of transcribing data to the card.

The cards used in the other illustrations accompanying this article are not exploratory to the same extent as the layout shown in figure 6, but for purposes of explaining the use of visual-analysis cards they will serve just as well. Only the items described earlier in this chapter will be noted. It can be seen readily, however, that a card prepared according to the sample layout would provide many more items for a study of relationships.

By way of bringing out the reliability of the patterns of characteristics as observed on the edges of sorted cards in a pack, the cards were equally divided in each size group and the halves assembled in two packs making two groups of cards sorted by size of farm (figure 7). With but slight variations in degree, exactly the same relationships appear in each half of this pack as were described in the explanation of figure 5. The size-of-farm pattern would, of course, be identical in both parts.

At this point it might be expected that all information had been obtained concerning relationships between size of farm and the items under observation. Except for the fact that each of the observations previously made can now be brought out in "bold relief" as it were, this is true. Not infrequently, however, in making an exploratory study, some of the observations are quite contrary to what was expected and the researcher desires further proof. If each of the relationships is approached through another avenue and similar results obtained, the added evidence is most welcome.

[illegible]

FIGURE 6.—Card designed primarily for visual-analysis purposes.

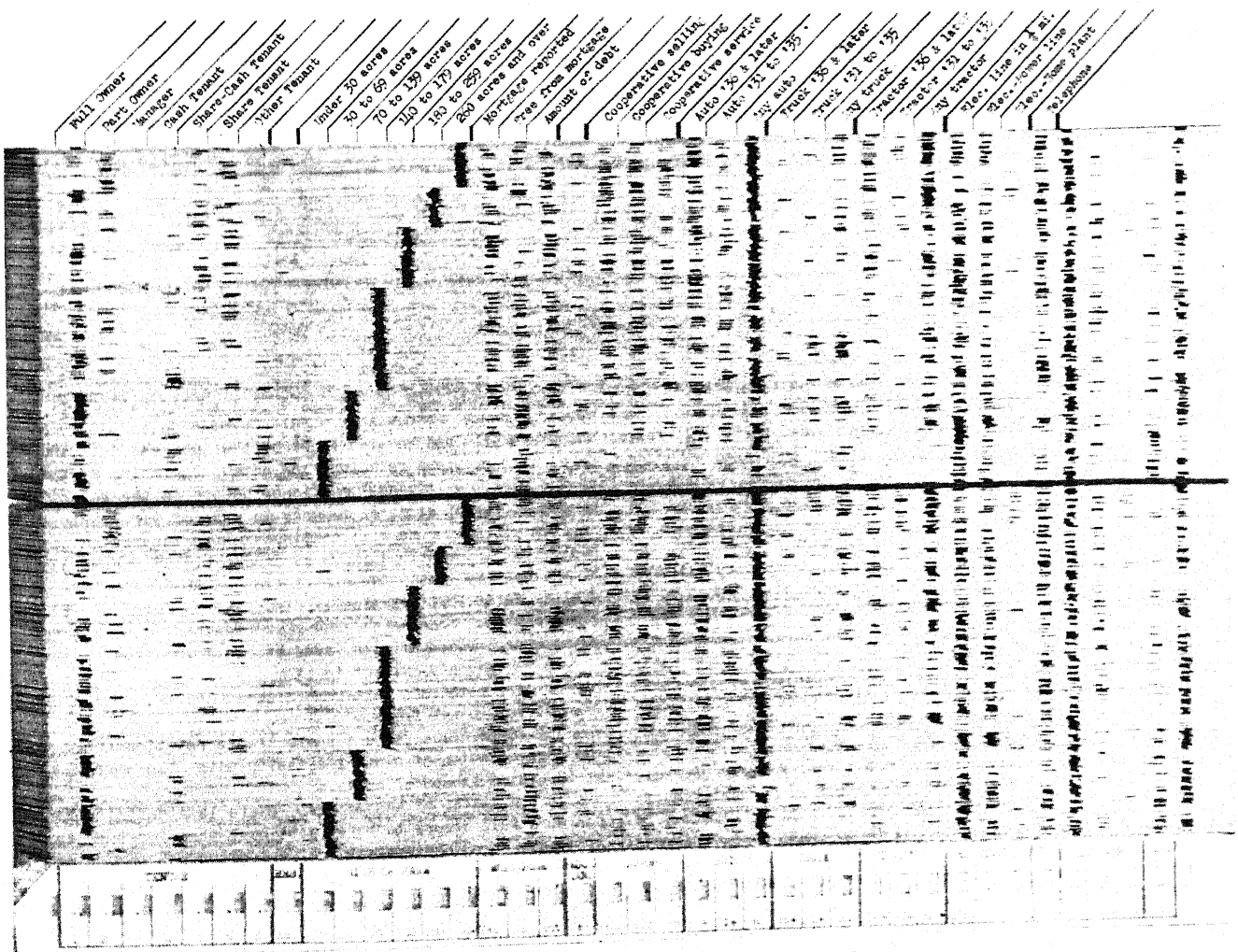


FIGURE 7.—Cards sorted first by size of farm, then each size group halved to illustrate the stability of the relationships visible.

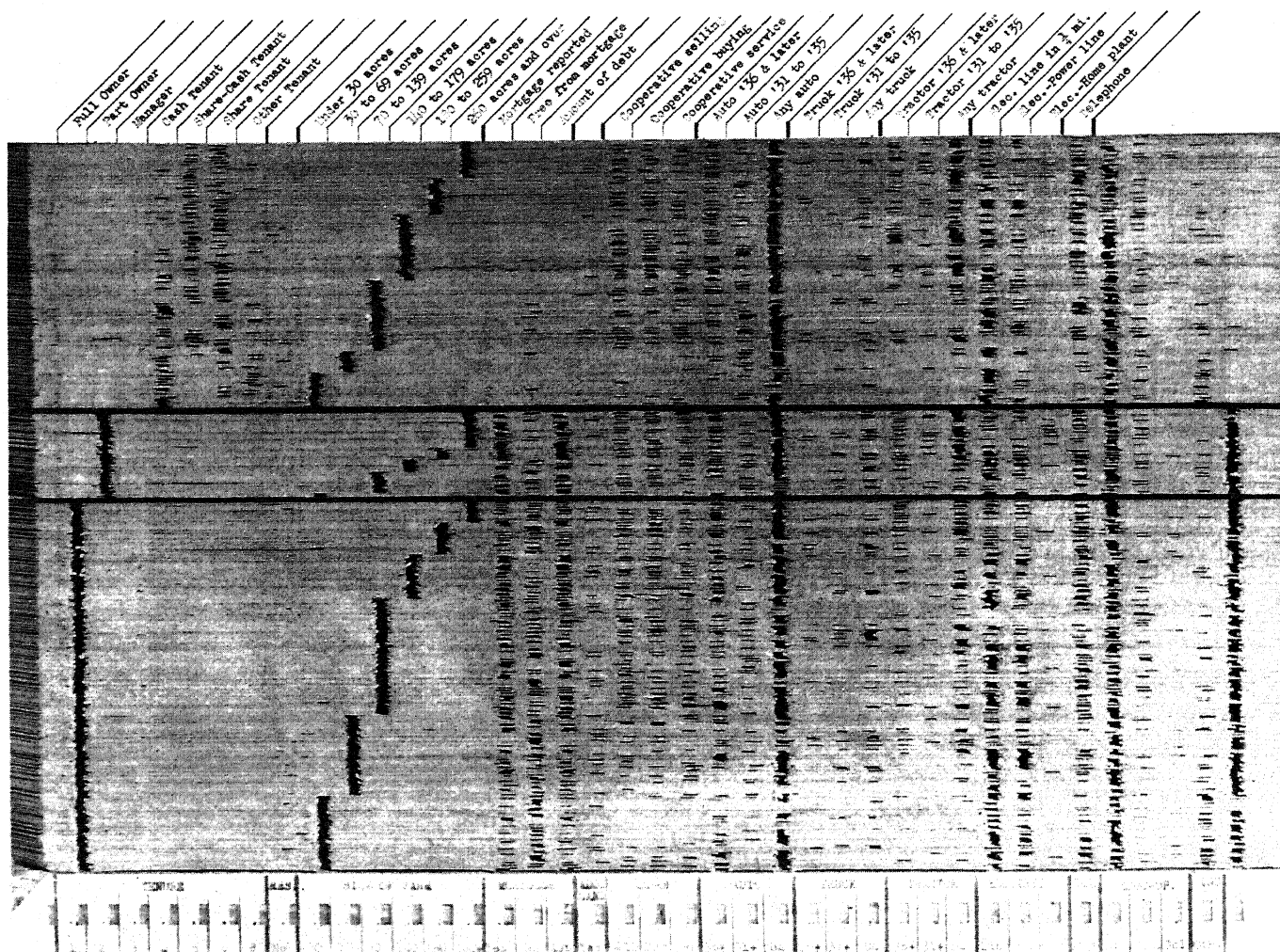


FIGURE 8.—Cards sorted to show the effect of tenure on size of farm; also, that some characteristics related to size of farm are not affected by tenure.

In figure 8 the cards have been sorted by tenure after the two halves of each size-of-farm group had been reassembled. It will be recalled that after the initial size sort, the full owners appeared to decrease as size of farm increased with the opposite trend for part owners. All of the full owners are now together in the lower section of the pack and all of the part owners in the center section. A direct comparison can be made of the size-of-farm patterns for each group. If the original observation is true, a comparison of the two patterns should show increasing proportions of each size group being represented in the part-owner section, starting with a relatively small portion in the smallest size group. This is exactly what is visible on the edges of the cards. The part-owner section shows only a very small proportion of the cards in the "Under 30" acre group, none in the "30 to 69" acre group, but above that point shows increasing proportions when compared with similar groups in the full-owner section.

In order to record this relationship for future reference (photographs such as these would be perfect if suitable equipment were easily available), the patterns of size of farm may be plotted. A simple and rapid method is to chart the different lengths of grooves in a manner that will show their changing relations. Figure 9 illustrates the changes taking place in the relative lengths of grooves as the farms increase in size.

A method requiring more time, but having the advantage of slightly greater accuracy, is an actual count of the cards in each group, and either plotting the values directly or expressed as a percentage of the total cards in each section.

At every sorting of the cards new concentrations of slots appear. Hence, no new sort should be made until all relationships are examined in the light of the current sort and compared with previous relationships noted. Beginning at the left of the pack, an item not mentioned before is the relationship of the different kinds of tenancy to size of farm. This is not an item in which we will continue to be interested in this paper, but is mentioned here as an example of the versatility of the system. Cash tenants represent the majority on the smaller farms, but practically disappear on farms of about 140 acres and up. Share-cash and share tenants occur in the reverse order, starting with farms of about 70 acres.

It is plainly evident in the mortgage block that a much higher proportion of part-owner farms are mortgaged than is true of full-owner farms.

The cooperative picture shows little relationship to tenure, the concentration of slots increasing with size of farm in each of the three groups of cards.

In the matter of automobiles, there is a definite tendency for more full owners to be without cars than for tenants. Considerably more than one-half of the group not reporting automobiles are full owners, while somewhat less than one-half of the group reporting automobiles are full owners. The next sort will be on automobiles and, if the sorting is carefully done in order to disturb the tenure sort as little as possible, this relationship will again be evident.

Trucks are reported most often on part-owner farms, and more often on full-owner farms than on tenant farms.

Tractors are reported more often on tenant and part-owner operated farms, which groups also have most of the larger farms.

Electricity seems available to full owners, part owners, and tenants alike, but is available more often to the smaller farms than to the larger farms.

The number of farms with telephones, again, increases with the size of farm.

In general, the same relationships are apparent as was true in the first summarization of relationships with some significant additions due to the tenure sort. The characteristics found to be related either directly or indirectly to tenure were size of farm, mortgage, automobiles, and, possibly, trucks and tractors.

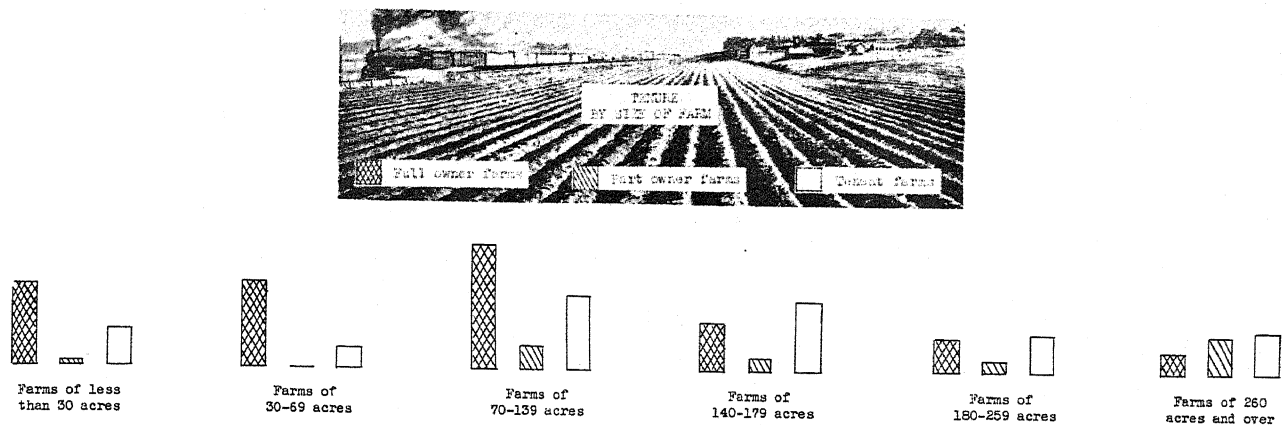


FIGURE 9.—Graphic illustration of the changes in number of farms by tenure which accompany changes in the size of farms used in these examples.

In figure 10 the cards are shown sorted by automobiles reported and not reported. Sorts on mortgages and cooperative business were omitted in the interest of brevity. However, as was evident in figure 8, these characteristics need not be lost to sight.

In this sort we again have the different groups (two in this sort) returned to the size-of-farm sort in order to compose the patterns. The contrasting patterns in figure 11 plainly show that as the size of farm increases, the number of farms without automobiles rapidly decreases.

After beginning at the left for a check on relationships, the statement made in a preceding paragraph concerning the relationship of automobiles to tenure can be checked. Of the farms not reporting automobiles (cards below the heavy divider card) nearly two-thirds are full-owner farms. A glance at figure 8 shows that approximately one-half of all farms, included in the study, are full-owner farms. Or better than this by the use of a divider, the accumulated lengths of grooves representing full owners in each case can readily be obtained. Comparison will then show that while approximately two-thirds of the farms not reporting automobiles are full-owner farms, less than one-half of the farms reporting automobiles are full-owner farms.

On the other hand, it has already been noted that the larger farms are operated for the most part by other than full owners. Hence, the apparent relationship between tenure and automobile ownership may, in reality, be a relationship between size of farm and automobile ownership or both may be influenced by items not shown on the card. It is highly probable, however, that characteristics having to do with size of operations would be prominent in the list, of which size of farm is as a rule one of the most significant. Were several such items "visible" on these cards, the proper relationship could easily be discovered.

A natural impulse in a situation of this kind is to bring all like reports together in order to roughly measure the influence of a given characteristic. For example, following the sort shown in figure 10, on observing the relationship between automobiles reported and size of farm, it might seem desirable to sort for each size-of-farm group, the reports and nonreports for automobiles and note the relative number of slotted and unslotted cards in each group. In many cases this will be found desirable as new relationships will appear associated

with the reporting (or nonreporting) of automobiles by size of farm. However, the number of groups of cards has been doubled and instead of a comparison between two size-of-farm patterns, the researcher must now compare six automobile patterns (one for each size-of-farm group). Added to this is the fact that owing to the fewness of cards in some of the groups, the observation would very likely be erratic.

In actual practice, it is usually found desirable when exploring the possibilities of the data to keep all or most of the cards in use in order that as many of the cards as possible be permitted to tell the story after each sort. In other words, if the primary purpose is to determine what items are related to size of farm, it is preferable to sort the cards each time in a manner to show the change made in the size-of-farm pattern. This has the advantage of keeping down the number of groups into which cards are sorted and also if the item on which the cards are sorted is not broken down into too many groups, the changes in the size-of-farm pattern may be plotted. On the whole, it systematizes the procedure. Any deviation from a specific procedure should be planned with a view to returning to the original with as little disturbance as possible to the previous sorts.

Returning to the discussion of relationships apparent after sorting the cards on automobiles, the following observations may be made concerning the farms represented by the cards used in these examples:

Fewer mortgaged farms are found among those not reporting automobiles.

Very little cooperative business is done from farms not reporting automobiles.

Trucks are reported no more frequently on farms without automobiles than on farms with automobiles.

The lack of desire for mechanization apparently carries over into tractors. Very few tractors are reported on farms without automobiles, even with size of farm given full consideration.

Availability of electricity decreases with increase in size of farm regardless of automobile ownership. But it is also quite obvious that electricity is available on a smaller proportion of farms not having automobiles than on those that do have.

Telephones seem somewhat scarcer on farms without automobiles, but are relatively more numerous on the larger farms in both groups.

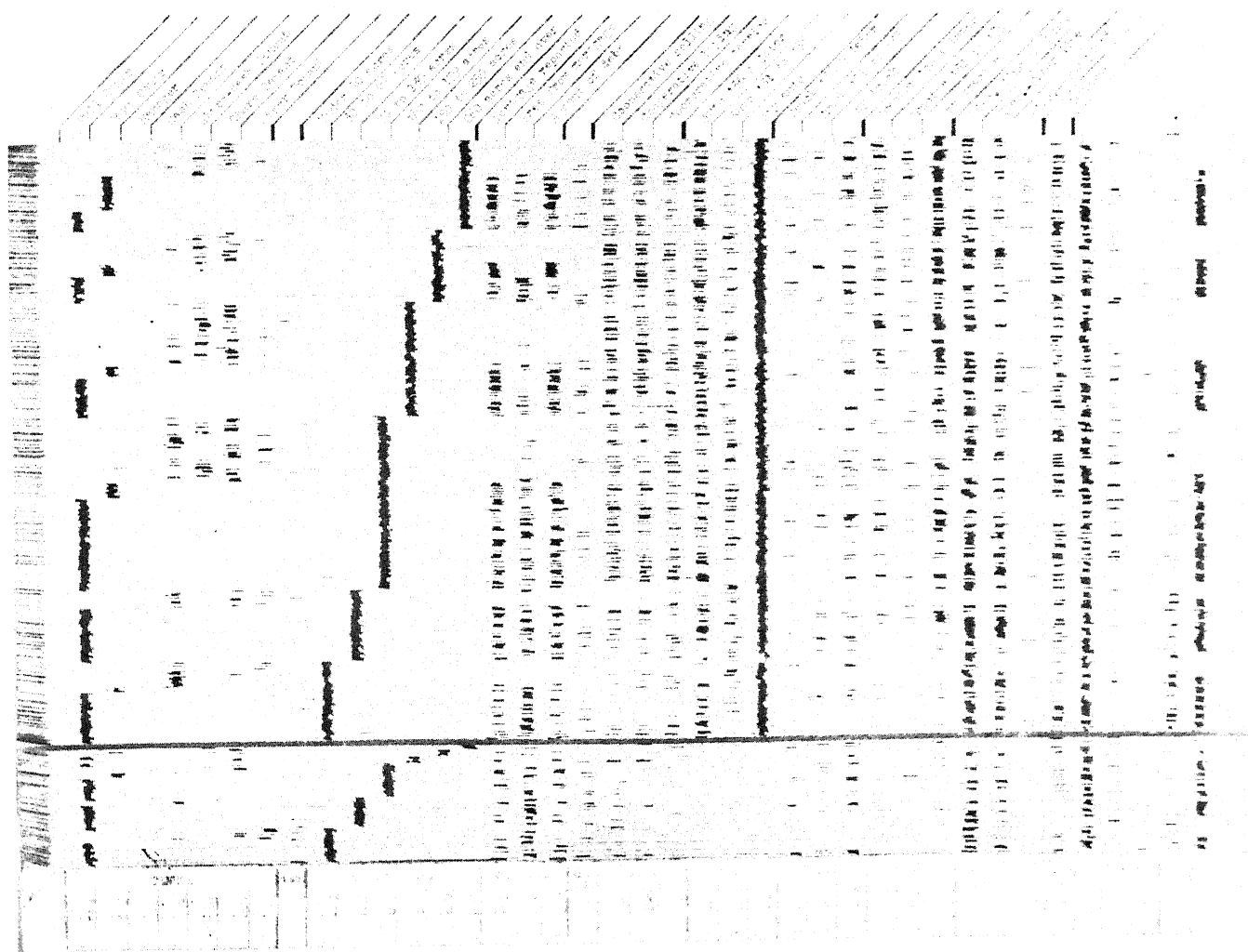


FIGURE 10.—Farms reporting automobiles compared with those not reporting automobiles. In these examples characteristics associated with large farms rarely appear on cards representing farms not reporting automobiles.

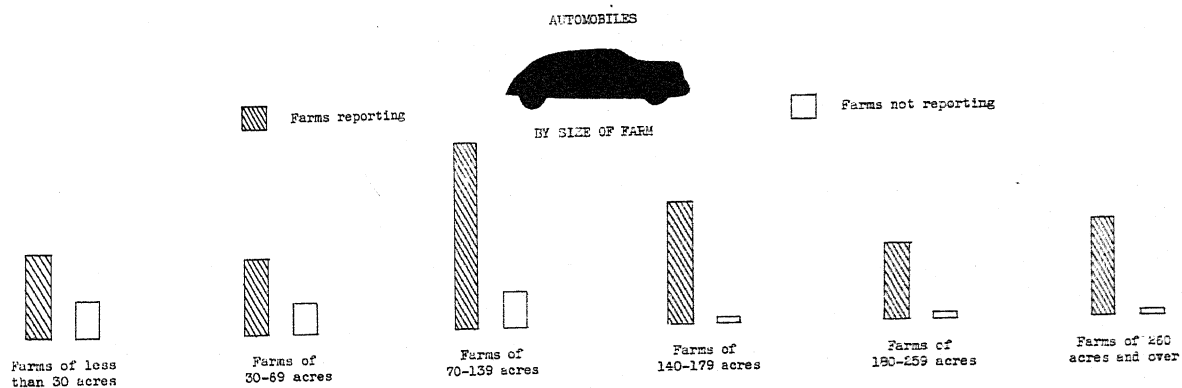


FIGURE 11.—Graphic illustration of the proportions shown in figure 10.

Of the items under observation, trucks have been less influenced by the different sorts than have the others. In figure 12 the cards are shown sorted according to whether or not a truck was reported; those indicating truck ownership are below the heavy line.

The chart shown in figure 13 indicates rather definitely that a greater proportion of farms in the larger sizes report trucks than do those in the smaller sizes.

A somewhat greater proportion of the farms not reporting trucks are mortgaged.

The distribution of farms doing business cooperatively is quite similar.

No significant difference is noticeable between automobile ownership in the two groups. Apparently about as many

operators reporting trucks report automobiles as do those not reporting trucks.

A somewhat greater number of truck owners report tractors than is the case with those not owning trucks.

The availability of electricity appears somewhat different in the two groups. If anything, this item increases in frequency with increase in size of farm for farms reporting trucks, while for farms not reporting trucks, the opposite is true. As a factor to be included in a study of plane of living of the farm operators represented by those cards, this item should be included.

The distribution of telephones is practically the same for both truck owners and nonowners.

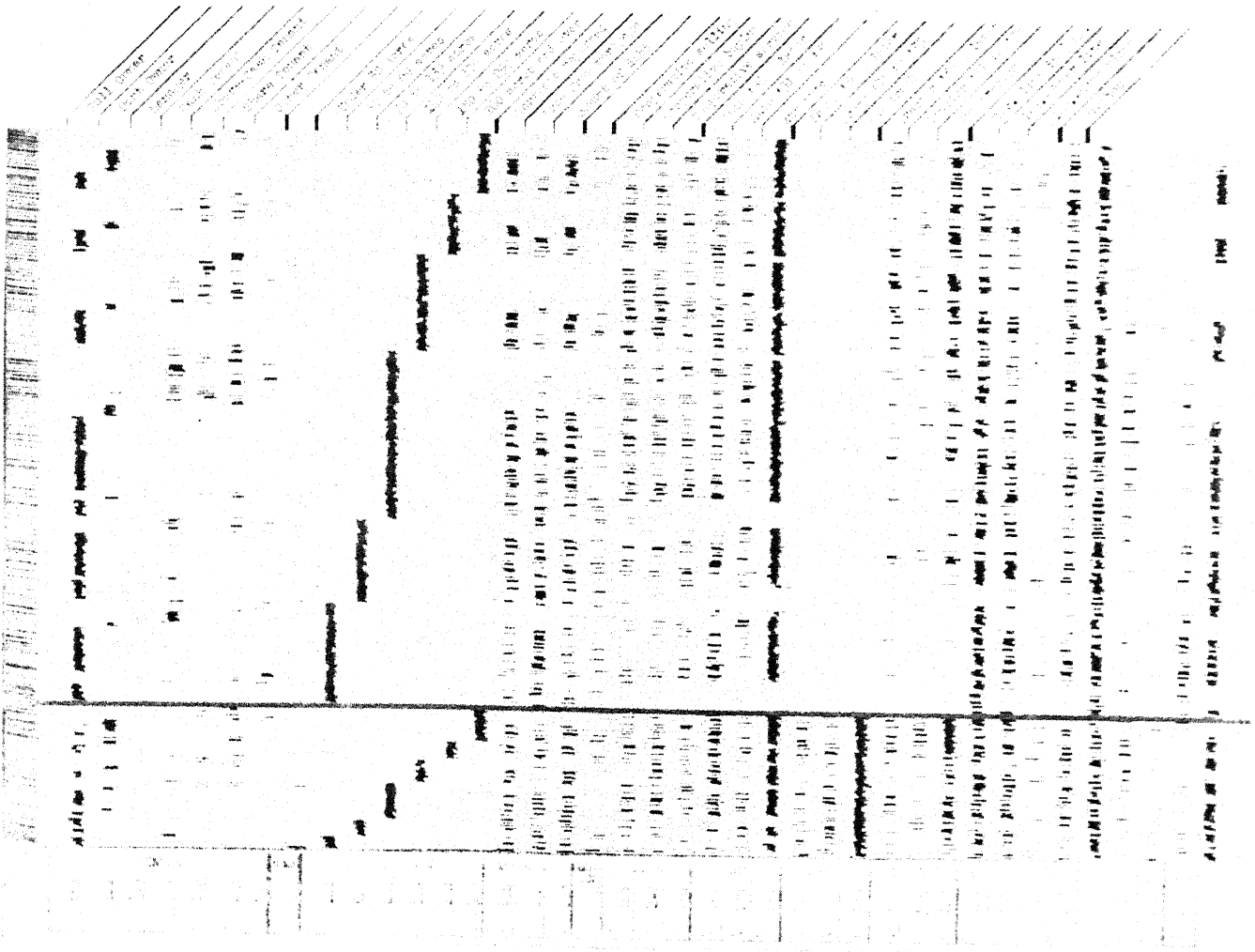


FIGURE 12.—Cards representing farms reporting trucks are shown below the heavy line.

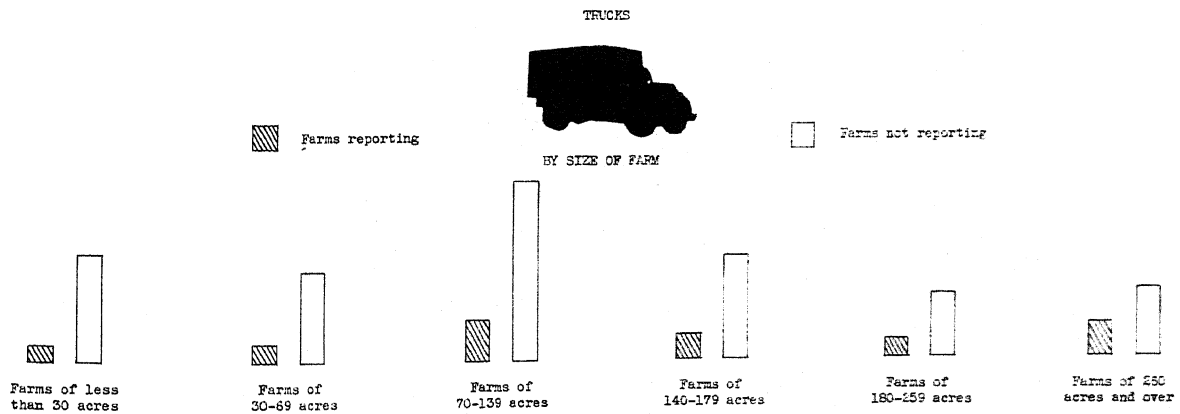


FIGURE 13.—Relative lengths of bars indicate the changing proportions in these examples of farms reporting trucks in six size-of-farm groups.

A sort of the cards by ownership of tractors (figure 14) shows a distinctly different size-of-farm pattern between tractor owners and nonowners. In figure 15 the respective patterns are in striking contrast.

Tractors are definitely associated with the size of farm.

A somewhat greater proportion of farms having tractors are mortgaged.

More tractor farmers do business through cooperatives than do nontractor farmers.

Practically all tractor farms have automobiles while many farms without tractors do not.

Tractor farms are more likely to have a truck than are nontractor farms.

Farms near an electric distribution line are more likely to have tractors.

Telephones are found more often on tractor-operated farms and seemingly with less regard to size of farm than is true of farms without tractors.

From the above observations, it is apparent that, among the operators represented by this sample, possession of a tractor is accompanied by a greater proportion of automobile and truck owners, and also operators with electricity available, than is true of operators without tractors.

The foregoing description of procedures is an attempt to clearly demonstrate the following:

1. By using an easily handled card or sheet, desired characteristics can be shown on the edges by assigning specific locations to each desired characteristic.

2. In this procedure slots are used to indicate the characteristics shown by the farm or operator in order that these same slots may be used in sorting the cards.
3. The slots on each card representing a farm form a distinctive pattern of characteristics for that farm.
4. When the cards are assembled in a pack and sorted into groups based on frequencies or bands, the grooves formed by the coincident slots form a pattern of that particular characteristic for the farms represented by the pack of cards.
5. Coincidence of slots representing characteristics other than the slot being used in sorting indicates relationships.
6. When a pack of cards is sorted according to any particular characteristic, the resulting concentrations of slots accompanying the primary sort furnish a pattern of characteristics for the farms represented.
7. When a pack of cards is sorted according to any particular characteristic, the greater the accompanying concentration of slots in any of the other characteristics, the closer the relationship between that characteristic and the one used in sorting. This is true providing, of course, that subsequent investigation does not disclose a variable that has been hidden up to that point.
8. Relationships are visible. Presence or absence of relationships is determined usually without tabulations other than occasional counting of the cards.

Figure 16 shows the cards sorted by progressive selection, a method which successively selects farms on the basis of specified characteristics. In this particular sorting the problem was to study the outstanding characteristics of "power" farms—those reporting possession of automobiles, trucks, and tractors, and also having electricity available.

The steps as far as shown in the illustration were predetermined, but it can readily be seen that by using a larger number of cards, particularly cards prepared in a manner to have more characteristics visible on the edges, a study of related characteristics would be rapid and accurate.

An outstanding feature of this method is the fact that an accumulation of characteristics is visible at all times. In any method it is usually necessary to do much trial and error grouping as the various related items are considered. By the use of visual-analysis cards, this exploratory grouping becomes a method of study with no chance of losing sight of any important item. No laborious labeling of different piles is necessary, neither is any great amount of space required. The groups of cards are self-labeled and any particular group may at any stage of the process be incorporated with another group. Negative groups may be investigated from time to time as a check on the direction the study is taking. As an illustration of this point, after determining a number of characteristics of power, it might be enlightening to compare these with the presence or absence of similar characteristics of farms not reporting some of the power items. This could be done quickly by running through the negative group in the same manner as the positive group. If some of the characteristics which were apparently related to power farms also appeared in significant strength in the nonpower group, this fact would be instantly visible. In addition, their comparative influences could be easily measured and the presence of intercorrelation could be determined.

With a large number of characteristics carried on the cards, the relationships and interrelationships become a fascinating study. For example, consider the occasional item which is related to the primary sort only when accompanied by a third item. A case in point would be farms reporting home plants for the production of electricity. Some of these farms have electricity available from a power line and some do not. It can be seen readily that the associations to be noted in the group reporting home plants with electricity available from a power line could be very different from those reporting home plants without the power line. The possession of a home plant when electricity was not available from a power line might be a characteristic associated with progressive farms, while the reporting of a home plant when electricity was available from a power

line might mean just the opposite. Such interrelationships are extremely difficult to keep in mind by hand grouping of reports or by mechanical sorting of cards, but by the use of visual-analysis cards they can hardly be lost to sight.

No attempt has been made in this article to evaluate the importance of visual analysis to statistical research or to any kind of research. To the researcher who has long struggled with the innumerable problems involved, some of the advantages of visual analysis will be immediately obvious. It is believed that actual use of the system will convince even the most skeptical that it is a worthwhile contribution to an important field of human endeavor.

The possibilities in the use of visual analysis have by no means been explored to the limit. On the contrary, the possible uses of the system are probably little realized. With anything so fundamental as a knowledge of relationships to proper understanding and evaluation in problems of research, any improvement in procedure that will contribute to this knowledge is a valuable addition. Many projects heretofore started on the basis of conjecture or, in many cases, pure guesswork, can now be carried through from the start on a sound basis.

The present article has been devoted to the use of visual analysis in purely statistical studies. This is by no means its only application. It is adapted to studies in the fields of sociology and economics, in summarizing and analyzing the results of experimental work of many kinds, and in personnel management, to mention just a few. Used in personnel studies, for example, with a card for each employee, the card provided with means of indicating by slots as many as possible of the characteristics associated with the employee and his job, it is obvious that there can be no practical limit to the amount of exact information that could be obtained. A sort of cards on the basis of age, for example, could well list the specifications of jobs best suited for the different age groups. Or sorted according to education, determine the value of education in the various jobs. Or, in the case of a large factory, all of the cards showing accidents could be segregated and the group compared with the no-accident group. Immediately, several factors of influence in accident prevention would be visible.

The cards shown in the illustrations in this chapter were used in connection with a special job done for the Meredith Publishing Company of Des Moines, Iowa. A card somewhat similar is being used in connection with special work for Magazine Marketing Service of New York City. Samples of the uses made of material assembled for these agencies are shown in chapter III, "Uses of Agriculture Statistics in Business and Research," of this publication.

